

# Il contributo di Antonino Mineo alla Statistica italiana

Marcello Chiodi

## Sommario

Antonino Mineo (1936-2008), professore ordinario di Statistica Metodologica, è stato una figura importante di statistico nel panorama italiano e internazionale. I suoi contributi scientifici sono stati prevalentemente nell'ambito della statistica multivariata, della statistica computazionale e delle distribuzioni di errori accidentali non normali. Fra le attività istituzionali è stato Preside della Facoltà di Economia di Palermo, e fondatore del Dipartimento di Scienze Statistiche e Matematiche Silvio Vianelli. In questo lavoro, che costituisce anche una bibliografia ragionata dell'opera di Mineo, vengo presentati i contributi principali di Mineo, inquadrandoli nella loro cornice storica

*Al mio maestro con affetto e stima*

## 1 *Brevi note biografiche e attività istituzionale*

Antonino Mineo si laurea il 10 novembre 1962 in Economia e Commercio con la votazione 110 e lode discutendo una tesi di Statistica Metodologica. Relatore è il Prof. Silvio Vianelli, di cui è allievo.

Prima assistente ordinario, poi libero docente, nel 1975 è vincitore del concorso a cattedra in Statistica. Dal 1975 ricopre la cattedra di Statistica Metodologica del corso di laurea in Scienze Statistiche ed Economiche della Facoltà di

Economia e tiene diversi corsi per supplenza fra cui Statistica 2 per il corso di laurea di Economia e Commercio, Teoria dei Campioni e Teoria e Tecnica dell'Elaborazione Automatica dei Dati per il corso di laurea in Scienze Statistiche ed Economiche.

Dal 1963 al 1978 è Capo Servizio Elaborazioni presso il Centro di Calcolo per la Ricerca Scientifica della Facoltà di Economia e Commercio, che dirigerà dal 1978.

Nel triennio 1980-83 è Presidente del Corso di Laurea in Scienze Statistiche ed Economiche. Nel triennio 1981-84 è Direttore dell'Istituto di Statistica della stessa Facoltà. Ha fatto parte del Collegio dei docenti del Dottorato di Ricerca in Statistica Computazionale, in consorzio con l'Università Federico II di Napoli.

E' Preside della Facoltà di Economia dal 1990 al 1996.

E' Direttore dell'Istituto di Statistica dal 1990 fino alla fondazione, nel 2002, del Dipartimento di Scienze Statistiche e Matematiche Silvio Vianelli, di cui è direttore fino al 2004.

### 1.1 *Attività scientifica nazionale e internazionale*

E' stato socio di diverse società scientifiche. In particolare nel quadriennio 1980-84 è componente del Consiglio Direttivo della Società Italiana di Statistica.

Nel quadriennio 1984-1988 è componente del Comitato regionale europeo della Bernoulli Society in rappresentanza dell'Italia.

Nel quadriennio 1986-1990 è componente il consiglio Direttivo dell'associazione scientifica *International Federation of Classification Societies* in rappresentanza dell'Italia.

Dal 1978 è membro eletto dell'I.S.I. (International Statistical Institute).

Nel 1982 ha organizzato a Palermo, per la prima volta in Italia, il 15° Congresso della Bernoulli Society for Mathematical Statistics and Probability, che vide la partecipazione di oltre duecento studiosi da tutto il mondo e in cui dimostrò grandissime capacità organizzative.

Ha tenuto più volte relazioni invitate ai convegni della Società Italiana di Statistica: i suoi interventi sono stati sempre caratterizzati da passione, spirito critico e spunti polemici.

Già colpito da una grave disabilità, nel 1999 tiene un seminario presso la Facoltà di Scienze Statistiche di Roma e nel 2000 organizza una sessione specializzata per la Riunione Scientifica della Società Italiana di Statistica a Firenze.

## *2 I contributi scientifici di Antonino Mineo*

Il contributo scientifico di Mineo si inserisce nel filone della scuola di Palermo fondata da Silvio Vianelli.

I contributi di Mineo alla statistica possono essere suddivisi in alcuni grandi filoni, affrontati da Mineo nel tempo a più riprese durante tutta la sua attività.

- la scomposizione dei miscugli di distribuzioni di frequenza che ha rappresentato uno dei primi contributi importanti di Mineo.;
- la scelta delle variabili nella regressione multipla e nell'analisi delle funzioni discriminanti
- Le tecniche di classificazione;
- Tecniche di analisi dei gruppi (cluster analysis);
- Errori non normali: curve normali d'ordine  $p$ ;
- tecniche di perequazione, distribuzioni ipernormali, superfici di risposta;
- contributi in ambito applicativo, fra cui: costruzione di scale di Strong, applicazioni in agraria, in ambito clinico, studi sui movimenti turistici.

Lavoratore instancabile e umile: non dava nulla per scontato durante lo studio ma, se si trattava di problemi applicativi, insisteva finché non trovava soluzioni ragionevoli e plausibili. Tenace non si arrendeva alle prime difficoltà o alle complicazioni portate da nuovi problemi di teoria o di calcolo.

Nella ricerca e nello studio era sempre molto aggiornato e invogliava gli altri ad esserlo, risultando attentissimo allo sviluppo delle strutture per la ricerca (attrezzature informatiche, biblioteche, etc.).

### *2.1 Il contributo di Mineo all'analisi multivariata in Italia*

Una parte del lavoro di Mineo può essere inquadrata in quella che negli anni si è delineata come *statistica multivariata*, per la quale Mineo ha fornito contributi per l'aspetto inferenziale, per quello più descrittivo-esplorativo ed anche per l'aspetto computazionale. Ha fornito contributi nell'inferenza in situazioni di non normalità ponendo attenzione agli aspetti numerici e sostanzialmente computazionali, per cui da un certo punto di vista Mineo può essere visto come uno dei precursori della cosiddetta *Statistica Computazionale*. Si è mosso sempre in un contesto essenzialmente parametrico e tutti i lavori metodologici, con rare eccezioni, sono sorti da problematiche reali. Mineo sottolineava la distinzione fra statistica e matematica, in quanto la prima fornisce i metodi per l'analisi di dati reali al fine di trarne informazioni sintetiche, mentre la matematica, in quanto logica formalizzata, studia in tutti i loro aspetti relazioni fra enti astratti (Minea, 1997).

Come in altri argomenti, l'approccio razionale e di natura parametrica a molti problemi, portava Mineo a tenere ben separati il contesto inferenziale, nel quale esiste un modello di riferimento ed occorre massimizzarne la portata informativa, da quello esplorativo, in cui sono i dati stessi a dovere suggerire il modello che più plausibilmente li descrive, secondo un principio di parsimonia e di semplificazione delle informazioni.

Minea non amava moltissimo la teoria isolata dai contesti applicativi: vedeva però spesso di un problema una soluzione magari prima per via intuitiva, puntando ad approcci semplici, parsimoniosi, e *interpretabili*, per poi formalizzare tali soluzioni in modo più rigoroso.

## 2.2 Scomposizione di miscugli di distribuzioni

Il problema della scomposizione dei miscugli di distribuzioni viene affrontato da Mineo nel 1964 (Minea, 1964a).

Occorre stimare i parametri di un miscuglio di due distribuzioni; per esempio per un miscuglio di due distribuzioni normali di densità:

$$f(x) = \frac{p}{\sigma_1 \sqrt{2\pi}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} + \frac{q}{\sigma_2 \sqrt{2\pi}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}$$

con

$$-\infty \leq x \leq \infty \quad 0 \leq p \leq 1 \quad p + q = 1$$

i parametri da stimare sono i due parametri medi  $\mu_1, \mu_2$ , le due varianze  $\sigma_1^2, \sigma_2^2$  e il parametro che indica la proporzione della prima popolazione,  $p$ .

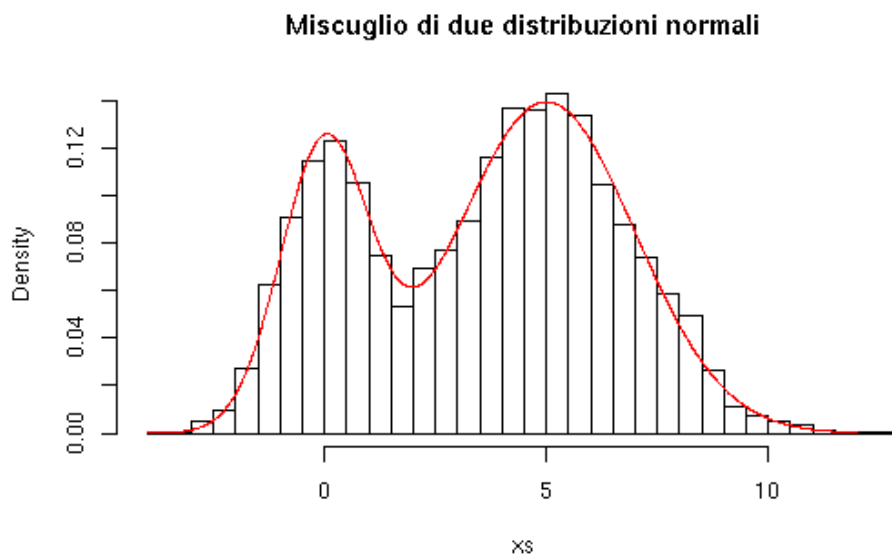


Figura 1: Miscuglio di due normali ( $\mu_1 = 0, \mu_2 = 5, \sigma_1^2 = 1, \sigma_2^2 = 4, p = 0.3$ )

Fino a quel momento il problema era stato affrontato per popolazioni con un solo parametro, o comunque quando i valori di alcuni parametri erano noti.

Pur essendo un problema di stima dalle forti implicazioni di tipo computazionale, Mineo lo affronta in modo ortodosso, considerando prima il problema della cosiddetta identificabilità dei parametri, ossia se una stessa sequenza di momenti teorici possa essere ottenuta da combinazioni diverse di parametri, e poi ponendosi il problema numerico della stima dei parametri. In particolare l'identificabilità dei miscugli di due distribuzioni normali risulta strettamente legato al fatto che la densità della distribuzione risultante sia bimodale e che si assumano note o le due medie o le due varianze (Teicher, 1960).

La stima dei parametri viene effettuata ricorrendo prima a stime iniziali fondate sul metodo dei momenti, e sulle mode della distribuzione empirica; successivamente vengono trovate *le stime di massima verosimiglianza*, risolvendo, con procedure numeriche iterative, un sistema di cinque equazioni in cinque incognite. Per risolvere questo sistema Mineo nel 1964 impiega il metodo delle sostituzioni successive, di convergenza lenta.

In un lavoro successivo (Minea, 1973) Mineo riaffronta il problema, sollecitato anche dal fatto che il tema è stato trattato nella letteratura scientifica negli anni '60, portando a nuove e più complesse problematiche. Mineo, programmatore instancabile, in questo lavoro si concentra sugli aspetti numerici e riporta anche il codice in FORTRAN II e le espressioni esatte degli elementi della matrice delle derivate seconde impiegate dall'algoritmo di Newton, cosa che consente a Mineo (cosa non tanto comune in quell'epoca, per problemi di tale complessità) di ricavare le matrici delle derivate seconde le cui inverse consentono un'approssimazione delle varianze campionarie degli stimatori.

### 2.3 *L'italiano è una lingua bellissima, ma inevitabilmente l'inglese è più letto...*

Fra i due lavori di Mineo sui miscugli, del 1964 e del 1973, esce su *Technometrics* nel 1966 un lavoro di Hasselblad (Hasselblad, 1966), che, senza citare Mineo, affronta numericamente il problema della stima dei parametri di un miscuglio di  $k$  popolazioni normali con il metodo della massima verosimiglianza.

L'approccio numerico è simile a quello di Mineo, ma l'impostazione metodologica e bibliografica era alquanto ridotta e non tratta affatto del problema della identificabilità degli elementi del miscuglio.

Confrontando gli approcci e le due bibliografie, la mia personale opinione è che Hasselblad si sia basato su altre fonti rispetto a quelle di Mineo, e che l'argomento ancora, per la sua complessità teorica e computazionale, non avesse allora una coerente e univoca sistematizzazione nella letteratura. I due lavori, pur trattando lo stesso argomento a due anni di distanza, hanno due sole citazioni in comune: Pearson (Pearson, 1894) e Hald (Hald, 1952) (dalla quale provengono i dati di una delle applicazioni riportate nel lavoro, ottenendo ovviamente gli stessi

risultati).

Il fatto che Hasselblad non citi Mineo probabilmente può essere spiegato anche col fatto che non conosceva il lavoro, scritto in italiano sugli Annali della Facoltà di Economia e Commercio di Palermo.

A suffragare questa congettura anche il fatto che in un lavoro su *Biometrika* Day (Day, 1969), affronta il problema della stima dei parametri di un miscuglio di  $k$  popolazioni normali multivariate. Day non cita nè Mineo nè Hasselblad ed ha una sola citazione in comune con Hasselblad e Mineo ossia il lavoro di Pearson del 1894!

Forse è questa la miglior dimostrazione del fatto che l'argomento ancora non fosse sistematizzato e che due riviste importanti, come *Biometrika* e *Technometrics* pubblicassero a cinque anni di distanza l'uno dall'altro lavori sullo stesso argomento con scarsa coincidenza bibliografica.

### 3 La scelta delle variabili

Mineo (Mineo, 1969) (Mineo, 1970) affronta il problema della scelta delle variabili nella regressione multipla e nell'analisi delle funzioni discriminanti (con  $k = 2$ ) con due approcci molto simili.

La scelta delle variabili rappresenta uno dei temi in cui si fonde l'interesse per diversi problemi statistici: per gli aspetti inferenziali della stima dei parametri, il problema della selezione a posteriori di un modello parsimonioso e l'importanza di implementare procedure di calcolo efficienti.

L'approccio è molto moderno e sta a cavallo fra le tecniche inferenziali e le tecniche esplorative; appartiene alla fase della ricerca nella quale si cerca il modello più adatto alla descrizione di una variabile dipendente.

La scelta viene effettuata partendo da un modello con tutte le variabili. Si procede quindi all'indietro, costruendo ogni volta dei test  $F$  in cui al denominatore viene posta una stima corretta della varianza, basata sul modello completo.

Mineo affronta nel 1969 il problema della scelta delle variabili nell'analisi discriminante (Mineo, 1969): sulla base di  $k$  campioni di osservazioni, ciascuno

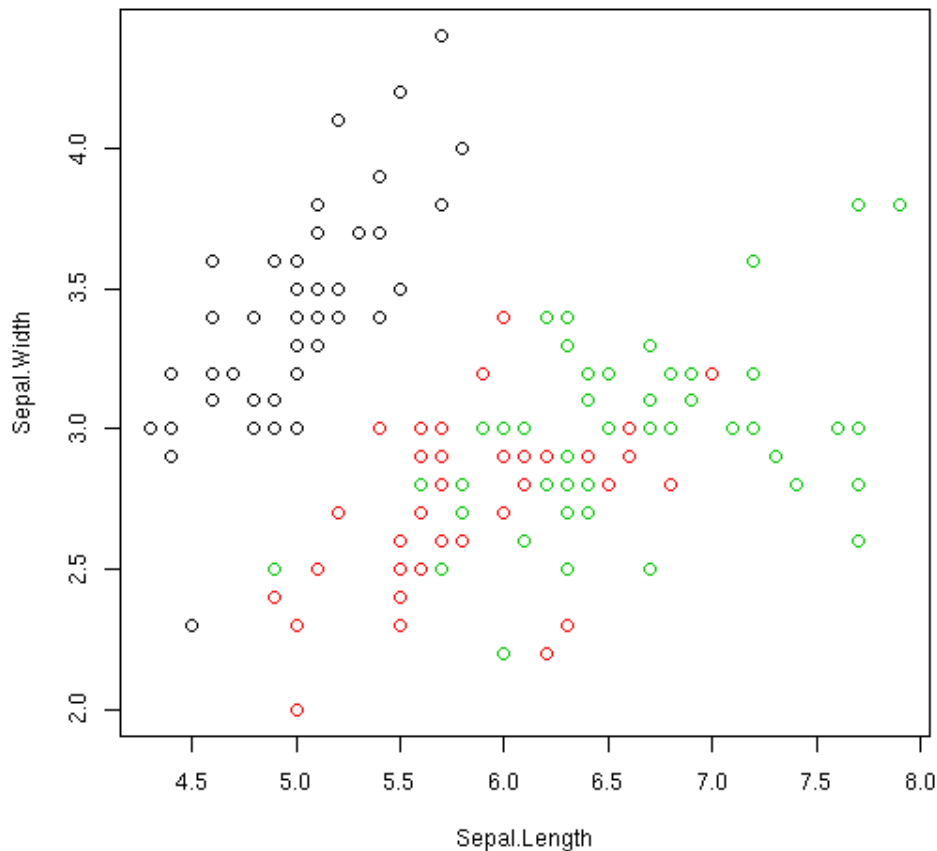


Figura 2: Discriminazione fra più popolazioni

proveniente da una diversa distribuzione normale multivariata, occorre trovare una *funzione discriminante* per assegnare una nuova osservazione ad una delle  $k$  popolazioni, minimizzando la probabilità di classificare in modo errato l'unità.

Come in analoghi problemi di previsione, l'aumento del numero di variabili non necessariamente migliora le capacità previsive della regola trovata. Mineo mostra con un esempio che la procedura proposta (tecnica di *selezione progressiva*) di selezione di variabili diminuisce la stima della probabilità di errata classificazione.

I lavori sull'argomento sono caratterizzati da un approccio pragmatico, in cui un contributo specifico è dato dall'ideazione di programmi di calcolo che



minimizzano il numero di operazioni necessarie per l'inversione delle matrici impiegate nelle varie fasi dell'algoritmo e dall'impiego di quantitativi minimi di memoria.

Dal punto di vista computazionale, Mineo si basa sul fatto che tutte le quantità necessarie al calcolo di  $R^2$  nella regressione multipla sono ricavabili dalla matrice di correlazione: occorre trovare un modo per evitare di calcolare tutte le regressioni possibili prima di arrivare ad una scelta di un particolare modello, cosa che avrebbe comportato la valutazione di  $2^{k-1}$  modelli distinti.

In particolare erano presenti nel codice degli accorgimenti atti a diminuire non solo i tempi di calcolo, ma fondamentalmente i requisiti di memoria richiesta all'elaboratore. Si ricordi che in quei tempi la memoria era misurabile in pochi Kilobytes, che dovevano bastare per il codice e per i dati!

Il metodo consentiva anche di ottenere una graduatoria delle variabili secondo la loro importanza, data dall'inversione dell'ordine di esclusione.

Un approccio analogo è stato poi adottato da Mineo per la scelta del numero delle componenti in una regressione polinomiale (Mineo, 1971). In questo modo Mineo giunge ad un polinomio con un numero minimo di termini, che può anche risultare incompleto.

#### *4 Tecniche di classificazione e raggruppamento*

Mineo affronta spesso nei suoi lavori tematiche di analisi statistica multivariata, variamente legate alle tecniche di classificazione. Questo è un termine molto ampio, sotto il quale, specie negli anni '60, '70 e '80, rientravano tecniche sostanzialmente diverse, che oltretutto non sempre corrispondevano nella letteratura di lingua anglosassone e in quella italiana.

Classification, clustering, grouping, clumping, discrimination, analisi dei gruppi, e così via, sono termini su cui Mineo ha scritto in due lavori di rassegna: *Per una classificazione della classificazione* (Mineo, 1978b) e *Criteri e metodi di classificazione statistica* (Mineo, 1986b); con il consueto approccio schematico e razionale, forniva una rassegna delle principali tecniche di classificazione e introduce, in accordo con le contemporanee tendenze più aggiornate della statistica multivariata italiana, *una schematizzazione che implicitamente o esplicitamente*

*ha avuto un forte seguito in Italia.* In quel lavoro in particolare Mineo distingue fra tecniche esplorative e non, in funzione della fase della ricerca nella quale ci si trova; ossia sulla base delle conoscenze che il ricercatore ha sul fenomeno in studio e sulle diverse informazioni sintetiche che egli vuole ottenere dall'analisi dei dati. Da questo punto di vista un problema primitivo è quello della ricerca delle strutture di variabilità dei dati. In questo ambito per esempio Mineo inserisce il problema della ricerca di gruppi omogenei al loro interno.

#### 4.1 *La cluster analysis*

Il problema della cluster analysis può essere così riassunto: vogliamo suddividere in  $k$  gruppi internamente omogenei  $n$  osservazioni su cui sono state rilevate  $p$  variabili quantitative.

Mineo pensa in fondo che la cluster analysis sia semplicemente un mezzo per potere distinguere delle forme in un insieme di  $n$  osservazioni a  $p$  variabili: quando  $p = 2, 3$  il nostro occhio distingue delle forme, dei gruppi, per lo meno può separare delle unità dalle altre o può giudicare se l'insieme appare omogeneo; con più di tre variabili ciò è impossibile e bisogna ricorrere a tecniche numeriche.

Mineo ritiene che fare questa suddivisione in modo esatto e univoco sia sostanzialmente impossibile, poichè troppe sono le scelte da fare in un contesto solo esplorativo: la scelta della metrica, della funzione obiettivo, del particolare algoritmo di ottimizzazione.

Anche in questo ambito, distingue fra gli aspetti computazionali e quelli di natura prettamente statistica: nella cluster analysis il problema computazionale è cruciale ed anche sostanziale, poichè è impossibile esaminare tutti i raggruppamenti di  $n$  osservazioni in  $k$  gruppi; per esempio con sole 20 osservazioni si hanno oltre 45 miliardi di possibili raggruppamenti in 4 gruppi, mentre il numero di possibili raggruppamenti di 50 oggetti in 4 gruppi è un numero di 29 cifre! Anche ammettendo di avere risolto il problema della metrica, della funzione obiettivo, della scelta delle variabili e del numero ottimale di gruppi, l'aspetto computazionale renderebbe il problema inaffrontabile in modo esatto.

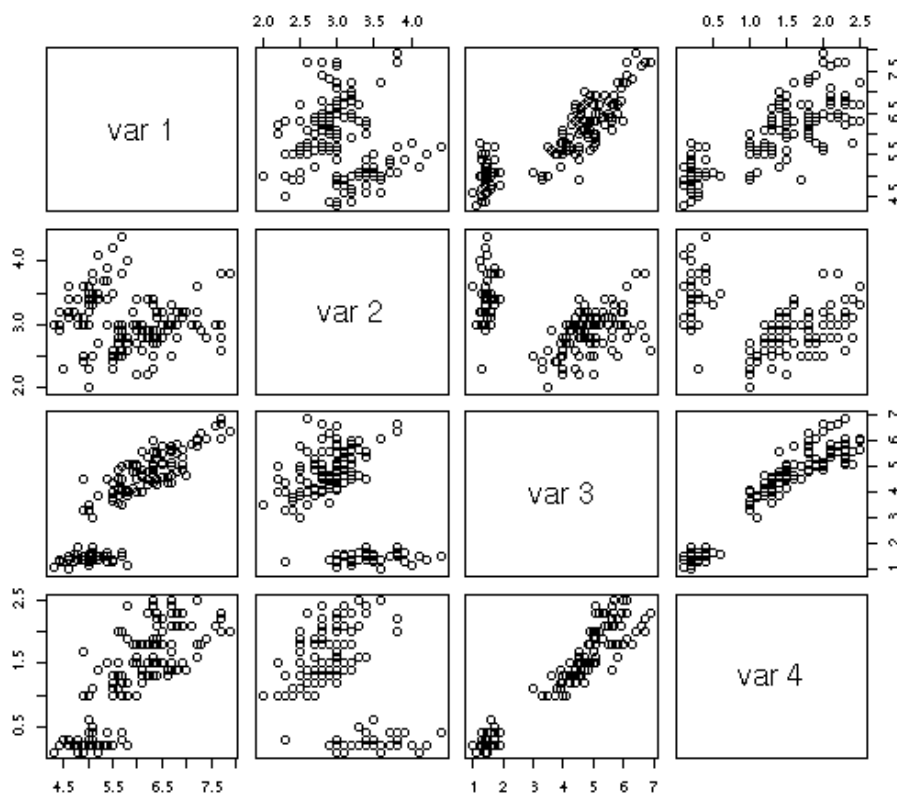


Figura 3: Identificazione di gruppi omogenei: cluster analysis

Mineo (Mineo, 1974) affronta la cluster analysis con il metodo del centroide. Come contributo originale al metodo che già si affermava negli anni '60, propone di partire da scelte ragionate della partizione iniziale. Per ciascuna variabile Mineo ottiene  $k$  coordinate, suddividendo l'intervallo di variazione in intervalli uguali:  $C_{1j}, C_{2j}, \dots, C_{kj}$  ( $j = 1, 2, \dots, p$ ) e costruisce una matrice di  $k$  centroidi a  $p$  componenti. Il metodo procede poi secondo l'algoritmo del centroide più vicino, includendo anche un criterio (non esatto) per la scelta del numero dei gruppi (Beale, 1969) che approssimativamente segue una distribuzione F.

Nell'applicazione proposta da Mineo nel lavoro, viene anche inserita una fase di scelta a posteriori delle variabili, scegliendo quelle che danno un maggior potere di determinazione di gruppi ben separati.

Successivamente (Mineo, 1985b), ritorna sul metodo precedentemente pro-

posto; trova, per ciascuna delle  $p$  variabili, una classificazione ottimale in  $k$  classi basandosi su un lavoro per la suddivisione ottima di un campione in  $k$  classi (metodo delle classi naturali, Mineo (1978a) descritto nella sezione successiva). Il criterio di scelta dei centri è stato perfezionato da Mineo in questo lavoro, con un'applicazione a dati astronomici, e reso più complesso, per evitare di ottenere una prima partizione con gruppi vuoti.

Notevole l'applicazione fatta nel 1985 (Minea, 1985a), in occasione di un convegno mirato all'analisi di casi di studio: gli stessi dati erano stati forniti a diversi studiosi e le varie soluzioni messe a confronto con successiva pubblicazione delle migliori. Fra queste quella di Mineo che affronta il problema della identificazione di particolari tipologie di crescita di un insieme di 30 ragazze seguite in un arco di 12 anni, e su ciascuna delle quali erano state rilevate 8 variabili. Mineo ottiene una ragionevole soluzione senza ricorrere a metodi fattoriali, cercando dei raggruppamenti per ciascun anno che consentissero di identificare delle strutture forti, resistenti nel tempo ed espresse nelle variabili originariamente osservate. Successivamente adatta delle curve logistiche alle medie dei gruppi trovati, per ottenere delle curve di crescita per soggetti omogenei.

## *5 Il metodo delle classi naturali*

E' con questo gruppo di lavori che Mineo ottiene visibilità internazionale (Minea, 1979), e da questo momento la sua attività scientifica si integra con l'attività istituzionale della Società Italiana di Statistica in campo internazionale. Culmine di questa attività internazionale sarà poi nel 1982 l'organizzazione a Palermo del 15° convegno della Bernoulli Society.

Quello sulle classi naturali (Minea, 1979) è il primo lavoro metodologico di Mineo pubblicato su una rivista internazionale (*Scandinavian Journal of Statistics*). Ancora una volta il nocciolo dell'idea è semplice, e riguarda aspetti computazionali e metodologici: è il primo lavoro in cui Mineo impiega esplicitamente tecniche di simulazione con metodo Montecarlo.

Il problema è banale e tutti ci siamo ritrovati a doverlo fronteggiare avendo davanti i valori di una variabile continua, rilevati con un certo grado di approssimazione. Lo scopo della suddivisione in classi, oltre che descrittivo ed esplorati-

vo, è anche il confronto fra la distribuzione empirica ed una distribuzione teorica per potere poi applicare un test  $X^2$  di bontà dell'adattamento.

Il lavoro ha un grande rilievo in Italia, ed ha avuto anche una buona visibilità internazionale. Si tratta probabilmente di uno dei primi lavori in Italia in cui lo strumento delle simulazioni viene impiegato in modo sistematico per approssimare risultati teorici, impossibili da ricavare con esattezza.

Sull'argomento sulla stessa rivista si ha uno scambio di osservazioni di Best e Rayner (Best and Rayner, 1979) e una risposta di Mineo che modifica leggermente le modalità di confronto con il metodo delle probabilità uguali (Mineo, 1981), ma mostrando sempre un grande vantaggio a favore del metodo delle classi naturali: anzi il lavoro di simulazione viene approfondito dai due Autori ed esteso.

E' noto che il test  $X^2$ , appropriato per la verifica che il campione provenga da una popolazione di una particolare famiglia, segue approssimativamente una distribuzione  $\chi^2$ , con  $k - m - 1$  gradi di libertà, essendo  $m$  il numero di parametri della distribuzione teorica da stimare. Nasce il problema se i parametri vadano stimati dai dati singoli originari, o dai dati del campione raggruppato: era noto che il metodo di stima influenzava la distribuzione campionaria del test  $X^2$ , ma Mineo lavora anche sull'influenza che ha il metodo di raggruppamento sulla distribuzione di  $X^2$ . In effetti tale influenza è duplice:

- il primo effetto è diretto, e agisce sulla suddivisione in classi e dunque sulle frequenze empiriche e teoriche che costituiscono  $X^2$ ; un altro effetto è indiretto ed è dato dal fatto che dal metodo di suddivisione in classi derivano le stime dei parametri basate sul campione raggruppato, sulla base dei quali si calcolano le frequenze teoriche.
- I metodi classici per la suddivisione in classi di un campione di osservazioni erano quello delle classi di ampiezza uguale e quello delle frequenze uguali: secondo Mineo questi due metodi, a prescindere dal numero di classi, non garantiscono che i valori (e conseguentemente le frequenze) si distribuiscano effettivamente casualmente attorno alle medie delle classi.

Il metodo proposto da Mineo consiste in un'aggregazione successiva dei valori in modo da minimizzare la devianza interna alle classi e conseguentemente massimizzare la devianza fra le classi. Mineo procede in modo ricorsivo: considera

inizialmente la classificazione del campione in  $n$  classi e classifica insieme le due osservazioni più vicine, ottenendo così una classificazione in  $n - 1$  classi. Si passa quindi ad una classificazione in  $n - 2$  classi aggregando le due unità (o classi) che forniscono il minor incremento di devianza interna. Si prosegue aggregando classi fino a giungere esattamente a  $K$  classi.

In questo modo si ottengono delle stime valutate sul campione raggruppato più vicine rispetto a quelle calcolate sui valori singoli: in particolare la stima della media risulta uguale, mentre la stima della varianza ha il minimo scostamento da quella di massima verosimiglianza valutata sui valori singoli.

Mineo approssima le distribuzioni campionarie delle tre quantità  $X^2$  calcolate con i tre diversi metodi di classificazione (classi naturali, classi di ampiezza costante e classi di frequenza costante) per mezzo di simulazioni condotte su campioni di numeri casuali provenienti da distribuzioni normali.

Per la generazione di tali numeri casuali, Mineo impiegò le tavole di numeri casuali della Rand Corporation (The Rand Corporation, 1955).

Fissati dei valori di  $n$  e  $k$ , su ciascuno dei campioni di ampiezza  $n$  vengono applicate le tre tecniche di classificazione e calcolati quindi tre valori del test  $X^2$ . Ripetendo il procedimento  $M$  volte, vengono ottenute tre sequenze di  $M$  valori simulati di  $X^2$  che costituiscono una approssimazione delle relative distribuzioni campionarie. Di tali distribuzioni Mineo calcola la media e la varianza e mostra come al variare di  $n$  e  $k$  il metodo delle classi naturali fornisca delle medie e delle varianze dei valori simulati di  $X^2$  molto vicini a quelli teorici; gli altri due metodi fornivano invece valori molto distanti da quelli teorici. Altri risultati teorici furono dimostrati da Chiodi (Chiodi, 1982).

## *6 Errori non normali: le curve normali di ordine $p$*

Mineo affronta il problema della stima dei parametri delle curve normali di ordine  $p$  già in un lavoro del 1980 (Mineo, 1980); il problema in sostanza consisteva nella stima dei parametri dai valori di un campione di osservazioni estratto da una

distribuzione di densità:

$$f(x) = \frac{1}{2p^{\frac{1}{p}} \Gamma\left(1 + \frac{1}{p}\right) \sigma_p} e^{-\frac{|x-\mu|^p}{p \sigma_p^p}}$$

In questo lavoro Mineo affronta anche il problema della stima dei parametri per campioni provenienti dalle *curve normali di ordine  $r$  per intervalli finiti*. Per la stima del parametro  $p$  introduce un particolare indice di curtosi,  $\beta_p$  il cui valore teorico è:  $\beta_p = p + 1$ ; ottiene quindi la stima di  $p$  eguagliando il valore empirico dell'indice a quello teorico, e per far questo occorre comunque risolvere un'equazione non lineare; gli altri due parametri  $(\mu, \sigma_p)$  vengono stimati dai momenti empirici.

Successivamente affronta il problema in un'ottica relativamente nuova e più moderna, stimolato anche dalla organizzazione a Palermo, come riunione satellite del convegno della Bernoulli Society nel 1982, di un corso sui metodi statistici robusti, organizzato da Hampel (Mineo, 1983b).

Mineo sosteneva che non è sensato applicare tecniche che modificano il contenuto informativo del campione, per garantirsi dell'eventualità che provengano da universi non normali, come per esempio le medie  $\alpha$ -trimmed, ed altri stimatori robusti della classe degli *M-estimators*: meglio cercare di adottare tecniche di stima che tengano conto del fatto che il modello di provenienza dei dati possa essere più generale di quello normale, per esempio il modello delle curve normali di ordine  $p$ , che rappresenta una famiglia di curve unimodali, simmetriche con curtosi variabile in funzione del parametro di struttura  $p$ .

Mineo inizia allora una serie di lavori (Mineo, 1983a) in cui i parametri di posizione e di scala vengono stimati mediante massima verosimiglianza mentre per il parametro  $p$  adotta tecniche diverse in generale fondate sulla curtosi campionaria. Questi stimatori sono stati messi a confronto con una batteria di stimatori *robusti* mediante tecniche di simulazione: Mineo mostra come le caratteristiche di tali stimatori siano sempre inferiori a quelle degli stimatori di massima verosimiglianza al variare dell'ampiezza campionaria e del vero valore del parametro  $p$  (Mineo, 1986a).

Successivamente applica tecniche simili alla regressione lineare semplice (Mineo, 1989). In effetti Mineo, anche nelle sue lezioni, dedicò molto ad inquadrare

nella giusta prospettiva storica il modello delle curve normali di ordine  $p$ , come curva generale per la descrizione della distribuzione di errori accidentali simmetrici (Mineo, 2002).

## *7 Il contributo di Mineo alla Statistica Computazionale*

E' uno degli aspetti più rilevanti dei contributi di Mineo: praticamente nessuno dei suoi lavori è privo di applicazioni, o di risultati ricavati per mezzo di simulazioni, anche nei lavori puramente teorici, in cui peraltro i calcoli venivano programmati ed effettuati dallo stesso Mineo.

Il termine *Statistica Computazionale* (Computational Statistics) si è affermato negli anni '70 e '80 e indica più che una serie di tecniche di calcolo, un approccio ai problemi della statistica, sia applicativi che teorici, inscindibile dall'uso del computer.

Mineo sfrutta appieno le potenzialità dello strumento di calcolo automatico fin da quando era caposervizio del Centro di Calcolo della Facoltà di Economia e Commercio di Palermo: programmava sempre in FORTRAN (solo in tempi recenti, e solo occasionalmente, ha programmato in BASIC) e la sua forza, che è stato in parte anche un limite, è stata quella di programmare sempre da solo tutto il codice necessario per un problema, dalla prima all'ultima riga. Questo da un lato lo porta a trascurare le potenzialità dei packages, ancora poco sviluppati ai tempi in cui inizia a lavorare, dall'altro lo porta ad un forte approfondimento delle tecniche numeriche migliori per la risoluzione approssimata di particolari problemi, fra cui:

- Ottimizzazione dei calcoli e della memoria per risoluzione di sistemi lineari e non lineari;
- Metodi di quadratura numerica e calcolo approssimato di probabilità integrali mediante approssimazioni a funzioni trascendenti;
- Tecniche di interpolazione numerica e costruzione di tavole relative per la costruzione di prontuari;
- Implementazione di studi di simulazione.



Notevole la sua capacità di programmazione in FORTRAN II: molti dei suoi lavori degli anni 60 e dei primi anni 70 sono completati in appendice con il codice utilizzato per ottenere i risultati, rara prova di trasparenza nei risultati, peraltro corretta dal momento che i problemi applicativi di cui si occupava Mineo mal si prestavano ad essere rieseguiti con carta e penna: l'unica possibilità era rendere pubblico il codice per far ripetere i calcoli su altre macchine.

Negli anni '80 e '90 poi, con la diffusione dei PC, passa anche all'uso del BASIC e superata la diffidenza iniziale verso questo linguaggio, inizia a impiegarlo, sempre con lo stile del vecchio programmatore cresciuto con le schede: non immissione di codice direttamente da tastiera e verifica diretta dei risultati, ma in generale scrittura su carta di tutto il codice e poi trascrizione su PC!

### *7.1 I prontuari delle probabilità integrali delle curve normali di ordine $r$*

Probabilmente il lavoro più oneroso dal punto di vista dei calcoli è rappresentato dai *Prontuari delle probabilità integrali delle curve normali di ordine  $r$*  (Minea, 1978c). In questo lavoro (in cui per la parte teorica è coautore Silvio Vianelli) Mineo dà grande prova delle proprie capacità di calcolo per un tipo di lavoro non propriamente da statistico, ma che rientrava pienamente nella grande tradizione *della scuola Palermitana di Vianelli*, che tante energie aveva profuso anche per la pubblicazione di manuali e prontuari.

Vianelli e Mineo con questo lavoro rendono accessibile a tutta la comunità scientifica il proprio bagaglio di conoscenze anche numeriche che possono essere utili per l'applicazione a problemi reali del modello delle curve normali di ordine  $r$ .

Il lavoro consiste nella tabulazione dei valori delle probabilità integrali per tale famiglia di curve, per diversi valori di  $r$  (da 1 a 3 ad intervalli di 0.05) e per valori della variabile  $x$  che procedevano per intervalli di 0.01. Le probabilità sono riportate con 7 cifre decimali, sebbene i calcoli siano stati effettuati con 10 cifre decimali esatte. Mineo riporta anche tutti i valori dei coefficienti necessari per l'applicazione di formule di interpolazione triangolare a 21 termini per potere calcolare per interpolazione con la stessa precisione numerica probabilità integrali

corrispondenti a valori di  $x$  ed  $r$  non tabulati.

Il lavoro è pieno di accorgimenti numerici volti a mantenere la precisione di calcolo in un range stabilito: si pensi che ancora nel 1978 l'uso delle calcolatrici da tavolo non era molto diffuso e si trattava comunque di strumenti costosi.

Il metodo impiegato da Mineo per il calcolo delle probabilità integrali si basa sulla nota proprietà che se  $k_r$  segue una distribuzione normale standardizzata di ordine  $r$ , allora  $|k_r^r/r|$  segue una distribuzione gamma di parametri  $(1/r, 1)$ . Per il calcolo di queste probabilità integrali Mineo impiega una integrazione di un classico sviluppo in serie a segni alterni per la quale risulta facile controllare il numero di termini necessari per conseguire l'approssimazione numerica desiderata.

## *8 Contributi sulla logica dell'inferenza*

Mineo alterna spesso, dagli anni 70 in poi, contributi specifici sia di natura metodologica che computazionale a contributi di carattere più generale, come descritto nelle pagine precedenti: scrive anche sulla genesi dei modelli di distribuzione degli errori di osservazione o su particolari approcci; porta anche una critica all'analisi delle corrispondenze (Minea and Chiodi, 1986), della quale criticava l'effettiva utilità ed anche il fatto che si trattava in fondo di una applicazione dell'analisi delle correlazioni canoniche.

Gli interessi di Mineo sono stati in effetti molteplici; fra gli altri ricordo un lavoro scritto agli inizi della carriera su una particolare distribuzione teorica leptokurtica (Minea, 1963), della quale ha riportato i parametri e ricavato proprietà teoriche. Un altro contributo ha riguardato le perequazioni meccaniche per le superfici di risposta (Minea, 1964b).

Numerosissimi poi i contributi di natura applicativa. Fra questi è da menzionare una serie di lavori condotti con Sprini (Minea, 1967b,a; Mineo and Sprini, 1967) sull'adattamento all'Italia del questionario sugli interessi professionali di Strong; poi molti contributi in ambito clinico, in ambito agrario, sull'autovalutazione degli esami, sulla dinamica dei movimenti turistici, sul sistema bancario, etc.

## 9 Conclusioni

In conclusione, Mineo è stato per la statistica italiana uno dei principali protagonisti, nei fatti, dello sviluppo delle tecniche computazionali e delle tecniche di statistica multivariata. In alcuni campi, come in quello dello studio delle distribuzioni non normali, il suo nome è stabilmente legato a quello delle curve normali di ordine  $p$ . Attento alle applicazioni, sostenitore del ruolo nettamente separato che hanno la matematica e la statistica, ha dato un apporto importante allo sviluppo ed al consolidamento della statistica nella realtà universitaria palermitana, grazie anche al suo impegno istituzionale, conscio del fatto che la statistica, a differenza di altre discipline, non vanta un radicamento secolare nella tradizione culturale e universitaria italiana.

Mineo ha messo nel suo lavoro e nel suo contributo scientifico tutto il suo modo di essere, dedicandovisi con passione: ha messo la tenacia, la capacità, il senso pratico e l'allegria che lo caratterizzavano, fornendo contributi che hanno lasciato il segno a livello scientifico e a livello umano.

## Riferimenti bibliografici

- Beale, E. M. (1969). Euclidean cluster analysis. *Bulletin del l'Institut International de Statistique*.
- Best, A. and Rayner (1979). A new grouping method for right evaluation of the chi-square test of goodness of fit. *Scandinavian Journal of Statistics*, (6), 145–153.
- Chiodi, M. (1982). Di alcune proprietà ottimali del metodo di raggruppamento delle classi naturali. *Alcuni lavori di analisi statistica multivariata. SIS. a cura di Renato Leoni*, pages 25–40.
- Day, N. E. (1969). Estimating the components of a mixture of normal distributions. *Biometrika*, **56**(3), 463–474.
- Hald, A. (1952). *Statistical theory with engineering application*. Wiley and sons, New York.

- Hasselblad, V. (1966). Estimation of parameters for a mixture of normal distributions. *Technometrics*, **8**, 431–444.
- Mineo, A. (1963). Di una particolare distribuzione statistica ipernormale. *Annali della Facoltà di Economia e Commercio dell'Università di Palermo*, **XVII**(2), 259–272.
- Mineo, A. (1964a). Sulla scomposizione dei miscugli di distribuzioni di frequenza. *Annali della Facoltà di Economia e Commercio dell'Università di Palermo*, **XVIII**(2), 121–145.
- Mineo, A. (1964b). Sulle cosiddette perequazioni meccaniche delle superfici di frequenza e di risposta. *Annali della Facoltà di Economia e Commercio dell'Università di Palermo*, **XX**(1), 229–243.
- Mineo, A. (1967a). Analisi critica dei metodi statistici impiegati nella costruzione delle scale di interessi professionali di Strong. *Annali della Facoltà di Economia e Commercio dell'Università di Palermo*, **Anno XXI**(1).
- Mineo, A. (1967b). Sulla attendibilità del test di interessi professionali di Strong. *Atti della XXV Riunione Scientifica della Società Italiana di Statistica, Bologna*, pages 1105–1112.
- Mineo, A. (1969). Su un nuovo criterio ottimo per la scelta delle variabili delle funzioni discriminanti. *Atti della XXVI Riunione Scientifica della Società Italiana di Statistica, Firenze*, (2), 379–404.
- Mineo, A. (1970). La scelta delle variabili nell'analisi della regressione multipla e delle funzioni discriminanti. *Statistica*, **XXX**(2), 223–255.
- Mineo, A. (1971). Sulla scelta del grado di una funzione di regressione polinomiale nell'analisi delle serie storiche. *Statistica*, **XXXI**(3), 529–554.
- Mineo, A. (1973). Ancora sulla scomposizione dei miscugli di distribuzione di frequenza. *Annali della Facoltà di Economia e Commercio dell'Università di Palermo*, **XXVII**(4), 231–260.

- Mineo, A. (1974). Un nuovo criterio per l'analisi dei gruppi - cluster analysis. *Annali della Facoltà di Economia e Commercio dell'Università di Palermo*, **XXVIII**(1), 3–36.
- Mineo, A. (1978a). On a new method for grouping. *1° meeting degli statistici europei aderenti alla Bernoulli Society*, page 149.
- Mineo, A. (1978b). Per una classificazione della classificazione. *Due temi di analisi statistica multivariata*, pages 233–245.
- Mineo, A. (1978c). Prontuari delle probabilità integrali delle curve normali di ordine  $r$ . *Grafica Studi Europa*, **II**, 39–58.
- Mineo, A. (1979). A new grouping method for right evaluation of the chi-square test of goodness of fit. *Scandinavian Journal of Statistics*, (6), 145–153.
- Mineo, A. (1980). La stima dei parametri delle curve normali di ordine  $r$  per intervalli infiniti e di quelle per intervalli finiti. *Studi in onore di Paolo Fortunati*, pages 565–577.
- Mineo, A. (1981). Rejoinder to best and Rayner's: A note on Mineo's grouping method for the chi-square test of goodness of fit. *Scandinavian Journal of Statistics*, (8), 187–188.
- Mineo, A. (1983a). The best combination of observation to estimate location and scale parameters. *44-ma Sessione dell'Istituto Internazionale di Statistica*.
- Mineo, A. (1983b). La migliore combinazione delle osservazioni per la stima dei parametri di intensità e di dispersione. *Relazione invitata di statistica metodologica alla XXXI Riunione Scientifica della Società Italiana di Statistica*, pages 463–487.
- Mineo, A. (1985a). Multivariate and longitudinal data on growing children: solution using clustering method. *Data analysis: the Ins and Outs of solving real problems*, Plenum Press, pages 67–87.
- Mineo, A. (1985b). A new criterion for the choice of seed points for a nearest centroid cluster analysis. *Rivista di Statistica Applicata*, **18**, 191–198.

- Mineo, A. (1986a). La migliore combinazione delle osservazioni per la stima dei parametri di intensità e di dispersione. *Studi in onore di Silvio Vianelli*, **I**, 519–548.
- Mineo, A. (1986b). Problemi e metodi di classificazione. *Riunione della società Italiana di Statistica*, (1), 83–108.
- Mineo, A. (1989). The norm- $p$  estimation of location, scale and simple linear regression parameters. *Lecture Notes in Statistics: Statistical Modelling*, (57), 222–233.
- Mineo, A. (1997). Il diverso contributo della statistica e della matematica al processo induttivo-deduttivo della conoscenza. *Induzioni*, **15**, 65–74.
- Mineo, A. (2002). Curve normali di ordine  $p$ . *Studi in onore di Angelo Zanella*, Milano: Vita e Pensiero, (503-521).
- Mineo, A. and Chiodi, M. (1986). L'analisi delle corrispondenze: un esame critico. *Riunione della società Italiana di Statistica*, pages 221–228.
- Mineo, A. and Sprini, G. (1967). L'adattamento italiano del questionario di interessi professionali di Strong: Svib. *Ingrassia*.
- Pearson, K. (1894). On the dissection of asymmetrical frequency curves. *Philos. Trans.*, **185**.
- Teicher, H. (1960). On the mixture of distributions. *Annals of Mathematical Statistics*, **31**, 55–73.
- The Rand Corporation (1955). *A million random digits with 100,000 normal deviates*, The Free Press, Glencoe, Ill. edition.